DOCUMENT RESUME

ED 431 964 CE 078 996

| | |
|---|---|
| TITLE | Assessing Employee Performance. |
| PUB DATE | 1999-03-00 |
| NOTE | 33p.; In: Academy of Human Resource Development (AHRD) Conference Proceedings; see CE 078 963. |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | EDRS Price MF01 Plus Postage. PC Not Available from EDRS. |
| DESCRIPTORS | Adult Education; Case Studies; Competence; Educational Needs; *Evaluation Methods; Interrater Reliability; *Job Performance; Job Training; *Labor Force Development; Needs Assessment; Office Occupations; *Performance Based Assessment; *Personnel Evaluation; Predictor Variables; Public Agencies; Statewide Planning; Systems Approach |
| IDENTIFIERS | Louisiana |

ABSTRACT

This document contains four symposium papers on assessing employee performance. In "Influence of Liking and Similarity on Multi-rater Proficiency Ratings of Managerial Competencies" (Reid A. Bates), the pattern of correlations identified between raters, independent variables, and different competencies suggests that raters may react differently to certain types of similarity and the nature of raters' reactions may vary with the type of skill or performance being rated. "A Cross-Case Study: Revisiting Implementations of the Performance Analysis for Training (PAT) Model" (Catherine M. Sleezer, Dale E. Kunneman) reports on five published case studies that describe systematic needs analysis based on the PAT model and reveal the influence of organization, decision maker, and analyst characteristics on the results yielded by the model. "Large-Scale Training Needs Assessment of State Government Office Support Workers: A Case Study" (Donna H. Redmann) discusses a case study of a statewide assessment of the training needs of office support and clerical workers in Louisiana state government. "Results Assessment System: From Evaluation to Assessment" (Richard A. Swanson, Elwood F. Holton III) describes the Results Assessment System, which is grounded in human resource development theory and practice. It is shown to be credible and practical. (MN)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 431 964

32 **Assessing Employee Performance**

**Chair: Paul Brauchle, Illinois State University**

32-1 Influence of Liking and Similarity on Multi-rater Proficiency Ratings of Managerial Competencies

Reid A. Bates, Louisiana State University

32-2 A Cross-Case Study: Revisiting Implementations of the Performance Analysis for Training (PAT) Model

Catherine M. Sleezer, Oklahoma State University

Dale E. Kunneman, Meridian Technology Center

32-3 Large-Scale Training Needs Assessment of State Government Office Support Workers: A Case Study

Donna H. Redmann, Louisiana State University

32-4 Results Assessment System: From Evaluation to Assessment

Richard A. Swanson, University of Minnesota

Elwood F. Holton III, Louisiana State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

R Bates D Redmann
C Sleezer R Swanson

2A

CE 072 956

# Influence of Liking and Similarity on Multi-rater Proficiency Ratings of Managerial Competencies

*Reid A. Bates*
*Louisiana State University*

*This study explored the predictive ability of liking and three forms of similarity for multi-rater proficiency ratings of managerial competencies. Liking and perceived similarity emerged as the strongest predictors of ratings for all raters. The pattern of correlations between raters, independent variables, and different competencies suggested there may be differences in rater reactions to certain types of similarity and that the nature of these reactions may vary with type of skill or performance that is rated.*

*Keywords: Multi-Rater, Proficiency Rating, Liking, Similarity*

In recent years, organizations have increasingly turned to multi-rater appraisal as a means to better observe and measure job performance, increase perceived fairness, reliability, and ratee acceptance (Harris & Schaubroeck, 1988; Latham & Wexley, 1982), and buttress the legal defensibility of appraisals (Bernardin & Beatty, 1984). Multi-rater appraisal instruments, sometimes referred to as 360° feedback instruments, have also gained popularity as developmental tools (Noe, Wilk, Mullen, & Wanek, 1997). Multi-rater appraisals provide individuals with information about their job-related knowledge and expertise from several sources including themselves, their supervisors, peers, subordinates, and customers. As developmental tools, such appraisals 1) broaden the evaluative perspective on an individual by using multiple data sources; 2) complement organizational trends toward empowerment and teamwork; 3) promote individual accountability for development; 4) enhance self-awareness and self-objectivity (Tornow, 1993); 5) improve communication between supervisors and subordinates; 6) and are typically inexpensive and simple to use (Borman, 1991).

Some research suggests the use of multi-rater instruments and the feedback they generate can lead to significant improvements in management skills over time (e.g., Hazucha, Hexlett, & Schneider, 1993). In order to understand how multi-rater instruments produce changes in managerial skills, Hellervik, Hazucha, and Schneider (1992) proposed a model that outlines the crucial steps of the behavior change process. The steps include 1) accurately assessing needs; 2) assigning behavioral standards; 3) formulating intentions to reach these standards; 4) learning new knowledge, skills, and abilities; 5) generalizing and maintaining new behaviors in the work environment. These scholars note that, in order to maximize the potential for behavior change, critical requirements must be met at each stage in the model. Because of its inaugural position in the change process model, one of the most important requirements is that appraisal information be accurate and credible.

This study examines the relation between two interpersonal variables, perceived similarity and liking, on multi-rater proficiency ratings. The study of these variables is important because performance ratings in general are susceptible to both intentional and unintentional biases. Biased ratings can affect both the credibility and accuracy of appraisals and the developmental feedback resulting from those appraisals. When appraisal information is perceived as either inaccurate or not credible, the likelihood of its acceptance by the ratee, and its power to guide or motivate managerial development is limited. The increasing use of multi-rater appraisals as performance management tools underscores the importance of understanding how such appraisals may be influenced by factors other than performance.

## Similarity and Liking

Recent models of both the appraisal process (Dipboye, 1985) and leader-member exchange (LMX) processes (Ferris, Judge, Rowland, & Fitzgibbons, 1994) have recognized interpersonal variables as potentially powerful factors influencing ratings. One important such variable is interpersonal affect. Affect refers to the degree of interpersonal attraction or like and dislike present in a relationship (Liden, Wayne, & Stillwell, 1993). It is a process distinct from cognition and has been cites as *the* major component of interpersonal relationships (Zajonc, 1980). Research has demonstrated that interpersonal liking can influence performance ratings in a direct and non-inferential manner (Alexander & Wilkins, 1982; Cardy & Dobbins, 1986); indirectly through the attributional processes of the rater (Dobbins & Russell, 1986; Mitchell, Green, & Wood, 1980); and through a schematic effect on the cognitive processing of performance information (Wayne & Ferris, 1990).

*Copyright © 1999 Reid A. Bates*

In addition, some research suggests that the degree of interpersonal attraction or liking may be influenced by the extent to which individuals perceive they are similar to others. That is, similarity implies a general agreement in views which people find rewarding (Byrne, 1971). Thus an individual perceived as similar is more attractive (and therefore more liked) and more apt to get positively biased performance ratings.

Recent research has begun to think about and assess similarity in three conceptually distinct ways. One conceptualization, demographic similarity (Turban and Jones, 1988; Zalesny & Kirsch, 1989), reflects the actual similarity of individuals along dimensions such as education level, gender, race, job status, job tenure, and so on. Although some researchers consider these factors unimportant to work relationships (Dienesch & Liden, 1986), others theorize they are critical to understanding organizational behavior (Schneider, 1987). From a cognitive processing perspective, demographic similarity can influence ratings because raters may use different schemas to classify information from similar and dissimilar ratees (Liden, Wayne, & Stillwell, 1993). If performance information is categorized on the basis of age, education, or gender similarity rather than actual performance, this may result in biased ratings. However, research examining the influence of demographic similarity on performance ratings is mixed. Some studies suggest a significant effect (Schmidt & Lapkin, 1980; Turban & Jones, 1988) whereas others indicate no effect (Cleveland & Landy, 1981; Wexley & Pulakos, 1983).

Another conceptualization of similarity, perceptual congruence, has been operationalized as the difference between an individual's self-description and other's descriptions of him/her (Wexley, Alexander, Greenwalt, & Couch, 1980); perceptions of communication frequency (Hatfield & Huseman, 1982); or behavior needed to obtain a merit pay increase (Turban & Jones, 1988). The limited research examining perceptual congruence has found higher levels of congruence related to job satisfaction and performance ratings. No research has examined the relationship between the perceptual congruence of required job competencies and performance ratings in multi-rater appraisals.

A final conceptualization of similarity refers to perceived similarity or similarity in attitudes, values, perspectives, and work habits between individuals. Research indicates perceived similarity can predict quality of LMX relationships (Liden et al., 1993), favorable treatment (Wexley & Nemeroff, 1974), and performance ratings (Pulakos & Wexley, 1983; Turban & Jones, 1988). Although little research has addressed this question, it has been suggested that perceived similarity, because of its attitudinal nature, may be the most potent factor influencing an evaluator's ratings of another individual (Schmitt, Pulakos, Nason, & Whitney, 1996).

Although research suggests liking and similarity may influence ratings, there is still little understanding of when and how these effects occur. The limited and often mixed results of previous research suggest, for example, the influence of similarity may vary by type of similarity, rater, or setting. Moreover, no research has examined the relationship between liking and different types of similarity in multi-rater appraisal settings. The present research examines the extent to which liking and three conceptually different forms of similarity can predict multi-rater proficiency ratings of managerial competencies. The following research questions were examined:

1. Are liking and the three types of similarity significant predictors of proficiency ratings resulting from a multi-rater appraisal?
2. Do liking or the types of similarity vary in their ability to predict proficiency ratings across different competencies?
3. Are there differences in the degree to which proficiency ratings from different rating sources (supervisors, peers, and subordinates) are associated with liking or similarity?

**Method**

Participants in this study were 142 mid-level managers in a state government agency. Positions included first-line supervisors (13%), second-line supervisors (12%), managers (57%), project leaders (5%), and other managerial positions (9%). Participants ranged in age from 29 to 66 with a mean age of 49. A little more than one-third of the participants (39%) had attained high school diplomas, 27% had some post-secondary training, 7% had associates degrees, 9% had bachelors degrees, and about 18% had graduate degrees. Twelve percent of participants had been in leadership positions in this organization for less than five years, about 30% between five and ten years, and the remaining 48% for more than ten years. The sample was predominantly female (64% female, 36% male).

***Procedure***

Subjects in this study were management and leadership personnel in a state government agency who had been selected to participate in a developmental assessment. The goal of the assessment was to identify competency strengths and weaknesses of management personnel as a first step in improving leadership and management in this agency. Participants completed a multi-rater managerial development instrument, the Leadership Effectiveness Inventory (LEI). A subset of target individuals (the ratees) rated themselves (n = 23), and were rated by immediate supervisors (n = 23), one or more peers (n = 44), and one or more subordinates (n = 53). Questionnaires were distributed during meetings held with the participants. The purpose of the meetings was to introduce the questionnaire, explain the developmental nature of the assessment, and to answer any questions participants may have had about the process or objectives of the assessment. Respondents were encouraged to complete the questionnaires within two days and return them to the researcher by mail.

***Measures***

All measures in this study were included in a questionnaire that employees were mandated to complete by the top administrator of this agency. Four independent variables were assessed. Raters responded to a measure of interpersonal liking and measures of three conceptually distinct forms of similarity. The liking scale adapted two items from previous research (Liden et al., 1993; Wayne & Ferris, 1990). These items asked "How much do you like this person as someone to work with" and "How much do you like this individual as a person?" Cronbach's alpha for this scale was .71.

Demographic similarity was measured in terms of discrepancies in gender, age, and educational levels between ratees and raters. Gender discrepancy was coded as the same (1) or different (0). Educational level was coded as high school diploma (1), post-secondary training (2), associate's degree (3), bachelor's degree (4), and graduate degree (5). Educational discrepancy was calculated as the absolute difference between rater and ratee. These scores were reversed so higher scores indicated greater similarity. Age was measured in years. Age discrepancy was calculated as the absolute difference between rater and ratee in years of age. Age discrepancy was transformed into similarity scores by dividing the difference scores by their respective standard deviation and reversing the resulting scores so that high scores indicated age similarity. All demographic scores were summed to yield a single score for demographic similarity.

A four-item scale adapted from Turban & Jones (1988) measured perceived similarity. Items included "This individual and I see work in much the same way"; "This individual and I are alike in an number of ways when it comes to work"; "This individual and I handle work problems in similar ways"; and "This individual and I are alike in terms of coming up with a similar solution for a work problem". The reliability estimate for this scale was acceptable (% = .93). Participants responded to these items along a five point Likert-type scale that ranged from 1 (not at all) to 5 (very much).

Perceptual congruence assessed the discrepancy between ratee and rater's perceptions of the proficiency levels required for technical competence (3 items) and interpersonal skills (5 items). For each competency item, respondents were asked to identify the proficiency required for successful performance in the ratee's job. Raters used a five point Likert-type scale that ranged from 1 (no proficiency) to 5 (very high proficiency). A sixth rating option (don't know) was also available but deleted from computation of congruence scores. Discrepancy scores were calculated as the absolute difference between rater's and ratee's required proficiency ratings. These scores were reversed so larger scores indicated greater congruence.

The dependent variables in this study were ratings of current proficiency for technical competence and interpersonal skills. Technical competence referred to an individual's ability to demonstrate technical proficiency, understand the nature of its impact in their areas of responsibility, and to appropriately apply policies and procedures. Interpersonal skills referred to proficiency in considering and responding appropriately to the needs, feelings, capabilities and interests of others, to provide feedback, and treat others equitably. These two competencies were assessed in order to explore whether certain types of competency ratings, such as those related to interpersonal skills, are more susceptible to affect or similarity influences than other, less person-oriented competencies (e.g., technical skills). For each competency item, respondents were asked to assess ratee current proficiency level using a five point Likert-type scale that ranged from 1 (no proficiency) to 5 (very high proficiency). A sixth rating option (don't know) was also available but was deleted from calculation of proficiency scores. Proficiency scores were combined to yield a single score, calculated as the sum of the ratings for each competency item.

The perceptual congruence and dependent variable measures were assessed using an adapted version the LEI, a leadership instrument designed to be used for management development in the public sector. The original LEI was developed by the U.S. Office of Personnel Management from a government-wide survey of over 12,000 respondents and was subsequently validated by teams of Federal executives, managers, and supervisors. The version used in this

study was modified slightly from the original version. Modifications, where necessary, were limited to changes in the wording of the items to better reflect job requirements and the culture of the agency. No items were added or deleted from the original instrument. This questionnaire required two ratings from respondents: current proficiency of the ratee and proficiency required for successful performance in the ratee's current job.

*Analysis*

Hierarchical regression analysis was used to examine how the variance in performance ratings was partitioned among predictors and to test the ability of the independent variables to explain the variance in the proficiency ratings. Independent variables were entered as shown in Table 1. Liking was entered first to determine how much variance the similarity variables could account for over and above this predictor. Of the similarity variables, demographic similarity variable was entered first because this represented actual as opposed to perceived or attitudinal similarities. Perceptual congruence was entered last, following perceived similarity, to determine how much variance this variable accounted for beyond the other similarity variables. Bivariate correlation analysis was used to investigate the question of differences in the associations of liking and similarity with different raters or on ratings with different types of competencies.

**Results**

***Research Question 1***

The dependent variable in this analysis was a single score representing the summed ratings for both technical competence and interpersonal skills. Hierarchical regression analysis results (see Table 1) indicated that:

1. Liking accounted for a significant proportion of the variance in proficiency ratings ($R^2$ = .48).
2. The addition of the demographic similarity variable did not increase the variance explained beyond that explained by liking ($R^2_{Change}$ = .00, ns).
3. The addition of perceived similarity to the model increased explained variance 9%, a significant increase (p = .001). The total variance explained by the model at this stage, with liking, demographic similarity, and perceived similarity as predictors was significant ($R^2$ = .57, p = .001).
4. The entry of perceptual congruence into the regression model increased total variance explained by a 5.4% (p = .001). Total variance in proficiency ratings accounted for by the full model equaled 62% ($R^2$ = .62, p = .001).

Overall, results indicated that liking together with the three types of similarity accounted for a significant proportion of the variance in proficiency ratings. Significant beta values (p = .001) were found for liking (ß = .40), perceived similarity (ß = .45) and perceptual congruence (ß = .27).

Table 1: Results of the Hierarchical Regression

| *Step* | *Predictor* | $R^2$ / *Adj* $R^2$ | $F_{model}$ *(df)* | □ $R^2$ | $F_{increase}$ *(df)* | □□ |
|---|---|---|---|---|---|---|
| 1 | Liking | .48 / .47 | 72.25** (1,80) | ---- | ---- | .40** |
| 2 | Demographic similarity | .48 / .47 | 36.25** (2,79) | .00 | .61 (1,79) | -.06 |
| 3 | Perceived similarity | .57 / .55 | 34.12** (3,78) | .09** | 16.06** (1,78) | .45** |
| 4 | Perceptual congruence | .62 / .60 | 31.64** (4,77) | .05** | 11.03** (1,77) | .27** |

□ Each successive model in the hierarchical regression analysis includes the variables previously entered.
□□ Beta values shown were calculated with all predictors entered in the regression equation.
* p □ .01 ** p □ .001

*Research Question 2*

Results in Table 1 indicate there are differences in the ability of conceptually different types of similarity to predict overall proficiency ratings. Liking in addition to both perceived similarity and perceptual congruence emerged as significant predictors while demographic similarity did not. Similar results were obtained in additional regression analyses examining the predictive power of these variables for the two separate components of the outcome variable (interpersonal skills and technical competence) (see Table 2). Liking, perceived similarity, and perceptual congruence all accounted for unique variance and were significant predictors of proficiency ratings for both competencies while demographic similarity was not. These results indicate there is no substantial difference in the predictive power of these variables across proficiency ratings of different types of competencies.

Table 2: Hierarchical regression analysis examining differential predictive power of liking and similarity types across competencies

| | | *Interpersonal Skills* | | | *Technical Competence* | | |
|---|---|---|---|---|---|---|---|
| *Step* | *Predictor* | $R^2$<br>*Adj* $R^2$ | $F_{increase}$<br>*(df)* | □-- | $R^2$<br>*Adj* $R^2$ | $F_{increase}$<br>*(df)* | □-- |
| 1 | Liking | .45<br>.45 | 66.46**<br>(1,80) | .36** | .35<br>.34 | 46.87*<br>*<br>(1,88) | .38** |
| 2 | Demographic similarity | .47<br>.46 | 2.42<br>(1,79) | -.12 | .35<br>.34 | .50<br>(1,87) | .07 |
| 3 | Perceived similarity | .57<br>.55 | 18.57**<br>(1,78) | .48** | .40<br>.38 | 6.78*<br>(1,86) | .35* |
| 4 | Perceptual congruence | .61<br>.59 | 8.35*<br>(4,77) | .21* | .45<br>.43 | 8.56*<br>(1,85) | .24* |

- Each successive model in the hierarchical regression analysis includes the variables previously entered.
-- Beta values shown were calculated with all predictors entered in the regression equation.
* p □ .01 ** p □ .001

*Research Question 3*

The question of whether there are differences in the degree to which developmental ratings from different rating sources (supervisors, peers, and subordinates) are associated with liking, perceived similarity, perceptual congruence, and demographic similarity was explored by bivariate correlation analysis. This analysis correlated ratings for interpersonal skills and technical competence from each rater with the independent variables. Results are shown in Table 3.

Table 3: Correlations between rating by source and independent variables for each competency rated

| | *Interpersonal Skills* | | | | *Technical Competence* | | | |
|---|---|---|---|---|---|---|---|---|
| | Liking | Similarity | Congruence | Demographic | Liking | Similarity | Congruence | Demographic |
| Supervisor | .56** | .88** | .47* | .01 | .47* | .58* | .26 | .28 |
| Peer | .60** | .67** | -.18 | .10 | .47** | .58** | .03 | .01 |
| Subordinate | .64** | .60** | .05 | -.29* | .55** | .48** | .13 | .03 |

All correlations are one-tailed
* p □ .05 ** p □ .01

Correlations are based on responses from 22 supervisors, 43 peers, and 25 subordinates. The resulting correlations ranged from .47 (p □ .05) to .64 (p □ .01) for liking, .48 (p □ .01) to .88 (p □ .01) for perceived similarity, -.18 (ns) to .47 (p □ .05) for perceptual congruence, and -.29 (p □ .05) to .27 (ns) for demographic similarity. These results indicate the relationship between both liking and perceived similarity and proficiency

ratings are consistent across source of rating and type of competency rated. The results for perceptual congruence and demographic similarity are less straightforward. Perceptual congruence was positively correlated with supervisor ratings of interpersonal skills but not with technical competence. In addition, it had no significant correlations with ratings from either peers or subordinates for either type of competency. Demographic similarity also had no significant positive correlations but showed a negative correlation with subordinate ratings of interpersonal skills ($r = -.29$, $p < .05$). Overall, these results suggest there may be differences in reactions to certain types of similarity and that the nature of these reactions may vary by rater/ratee relationship and type of skill or performance being rated.

**Conclusions and Discussion**

This study compared liking and three forms of similarity with multi-rater proficiency ratings of managerial competencies. Liking and perceived similarity emerged as the strongest predictors of ratings for all raters, explained unique variance in both interpersonal skills and technical competence, and were strongly correlated with all supervisor, subordinate, and peer ratings of both interpersonal skills and technical competence. The regression results and relatively high correlation ($r = .63$, $p < .001$) between these two variables support the similar-to-me association between liking and similarity. It is possible these two factors may combine to form one 'perceptual package' for raters appraising skills or performance. Results also support observations that perceived similarity may be the most likely type of similarity to influence ratings (Pulakos & Wexley, 1983). In general, these findings were not unexpected given previous research. However, this study did extend the findings to multi-rater proficiency ratings of qualitatively different kinds of competencies.

Although not as strong as either liking or perceived similarity, perceptual congruence was also a significant predictor of ratings and accounted for unique variance in the two competencies rated. This finding supports research cited earlier indicating that similarity in perceptions of job demands is associated with supervisor ratings. On the other hand, correlations between congruence, supervisor, peer, subordinate ratings, and the two competencies rated were mixed. Data showed that 1) perceptual congruence was associated with supervisor ratings of interpersonal skill but not for technical competence; and 2) it was not associated with peer or subordinate ratings for either competency. The findings indicate variations in how different rater react to congruence and suggests these variations may interact with type of skill rated. One possible explanation may be that supervisor-subordinate perceptual congruence about job requirements functions to increase task clarity. This allows subordinates to identify which job behaviors are most important (and most likely to be rewarded), and to focus their energies appropriately (Turban & Jones, 1988). The fact that perceptual congruence was associated with interpersonal skills and not technical competence could therefore reflect the relative importance of interpersonal skills in these managerial jobs. This interpretation further suggests the non-significant associations with peer and subordinate ratings indicate that the views of these individuals about what is important for job performance are not as critical as those of the supervisor in determining ratee behavior. On the other hand, one could expect perceptual congruence to play a larger in predicting ratings in team-oriented or other settings in which peers and subordinates are more involved with identifying and rewarding good performance.

Demographic similarity was not a predictor in any of the regression analyses although is did show a significant negative correlation with subordinate ratings of interpersonal skills. The inverse relationship between demographic similarity and ratings is perplexing. Although some research has found that older ratees are rated less favorably than younger ratees (Rosen, Jurdee, & Lunn, 1981), there was no significant age differences between subordinates and ratees to account for this finding. Nor were there significant differences in age or gender that might explain these findings. Moreover, it is difficult to speculate on the causes of this finding given current research and theory on the role of similarity perceptions in performance ratings. The absence of a similar relationship with technical competence suggests the possibility this finding may be an artifact of the data or a result of common method variance. In any case, these results add to the mix of findings with regard to demographic similarity and argue for further research directed at understanding if, how, and why this type of similarity influences performance judgments.

**Summary**

Little research has directly examined the interactions of potentially influential interpersonal variables on different raters and types of performance or skill being rated. The present research attempted to explore these issues. Findings

supported previous research identifying liking and perceived similarity as powerful predictors of performance ratings. Data also indicated that supervisor ratings are more closely associated with perceptual congruence than ratings of peers or subordinates. This, together with the pattern of correlations for both perceptual congruence and demographic, suggests (1) there may be differences in raters reactions to certain types of similarity; and (2) that the nature of these reactions may vary with type of skill or performance that is rated.

These findings are important when viewed in the context of multi-rater appraisals. Multi-rater appraisals are increasingly becoming a key component of an organization's performance management strategy (Church, 1998). The central rationale for the use of multi-rater appraisals is that job-related, behaviorally based performance information from multiple sources is a potent force for performance improvement. A key assumption underlying the application of these appraisals is that information from multiple sources will provide more valid and reliable information about an individual's job performance. However, the present research has shown that liking and different kinds of rater-ratee similarity can influence ratings and that this influence by vary by rater or rating task.

In practice, liking or similarity influences on the appraisal process may be difficult to control. For example, it is unlikely that a rater would ever encounter a rating situation in which his or her liking or perceptions of similarity were completed unrelated to ratees or ratings. Because interpersonal factors appear to be intertwined and inseparable from the rating process, it is important to understand that these factors are at work and how they may influence the appraisal process and the resulting information. The implication is that, to better manage the appraisal process, it may be more fruitful to approach raters as individuals in social roles whose feelings and interpersonal perceptions are significant influences on the rating process. In addition, the results of this and other research suggest the value of maximizing rater-ratee perceptual congruence about job requirements. In multi-rater appraisals, facilitating discussions between employees, supervisors, peers and subordinates with the explicit goal of verifying and reaching agreement on job requirements would enhance the credibility and utility of appraisal information. To the extent discrepant views on what is important for job performance have been reduced or eliminated, appraisal information will be better able to direct and motivate performance improvement efforts. Although enhancing perceptual congruence might not eliminate the influence of liking or perceived similarity in the appraisal process, it may help balance the effect of these factors on performance or skill ratings.

**References**


Alexander, E. R., & Wilkins, R. D. (1982). Performance rating validity: The relationship of objective and subjective measures of performance. Group and Organization Studies, 7, 485-496.

Bernardin, H. J., & Beatty, R. W. (1984). Performance appraisal: Assessing human behavior at work. Boston: Kent Publishing.

Borman, W. C. (1991). Job behavior, performance, and effectiveness. In M. D. Dunnette & L. Hough (Eds.), Handbook of industrial and organizational psychology (vol. 2, 2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Byrne, D. (1971). The attraction paradigm. New York: Academic Press.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: Liking as an integral dimension in evaluating performance. Journal of Applied Psychology, 71, 672-678.

Cardy, R. L., & Dobbins, G. H. (1994). Performance appraisal: Alternative perspectives. Cincinnati: Southwestern Publishing.

Church, A. H. (1998). Moderators of managerial behavior ratings: An analysis from multiple perspectives. In R. Torraco (Ed.), Proceedings of the Academy of Human Resource Development Annual Conference (pp. 747-754). Baton Rouge, LA: AHRD.

Cleveland, J. N., & Landy, F. J. (1981). The influence of rater and ratee age on two performance judgments. Personnel Psychology, 34, 19-29.

Dienesch, R. M., & Liden, R. C. (1986). Leader-member exchange model of leadership: a critique and further development. Academy of Management Review, 11, 618-634.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. Academy of Management Review, 10, 116-127.

Dobbins, G. H., & Russell, J. (1986). The biasing effects of subordinates likableness on leaders' attributions and corrective actions. Personnel Psychology, 39, 759-777.

Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: Test of a model. Organizational Behavior and Human Decision Processes, 58, 101-135.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, 41, 43-62.

Hatfield, J. D., & Huseman, R. C. (1982). Perceptual congruence about communication as related to satisfaction: Moderating effect of individual characteristics. Academy of Management Journal, 25, 349-358.

Hazucha, J. F., Hexlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. Human Resource Management, 32, 325-351.

Hellervik, L. W., Hazucha, J. F., & Schneider, R. J. ( 1992). Behavior change: Models, methods, and a review of evidence. In M. D. Dunnette & L. Hough (Eds.), Handbook of industrial and organizational psychology (Vol. 3, 2nd ed.). Palo Alto, CA: Consulting Psychologists Press.

Latham, G. P., & Wexley, K. N. (1982). Training approaches and a workshop to minimize rating error. In L. S. Baird, R. W. Beatty, & E. E. Schneir (Eds.), The performance appraisal sourcebook (pp. 85-90). Amherst, MA: Human Resource Development Press.

Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study on the early development of leader-member exchanges. Journal of Applied Psychology, 78(4), 662-674.

Mitchell, T., & Green, S., & Wood, R. (1981). An attribution model of leadership and the poor performing subordinate: Development and validation. In B. M. Staw & L. L. Cummings (Eds.), Research in organizational behavior, vol. 3. Greenwich, CT: JAI Press.

Noe, R. A., Wilk, S. L., Mullen, E. J., & Wanek, J. E. (1997). Employee development: Issues in construct definition and investigation of antecedents. In J. K. Ford, S. W. J. Kozlowski, K, Kraiger, E. Salas, & M. S. Teachout (Eds.), Improving training effectiveness in work organizations (pp. 153-189). Mahwey, NJ: Lawrence Erlbaum.

Pulakos, E. D., & Wexley, K. N. (1983). The relationship among perceptual similarity, sex, and performance ratings in manager-subordinate dyads. Academy of Management Journal, 26, 129-139.

Rosen, B, Jerdee, T. H., & Lunn, R. O. (1981). Effects of performance appraisal format, age, and performance level on retirement decisions. Journal of Applied Psychology, 66, 515-519.

Schmidt, N., & Lappin, M. (1980). Race and sex as determinant of the mean and variance of performance ratings. Journal of Applied Psychology, 65, 428-435.

Schmitt, N., Pulakos, E. D., Nason, E., & Whitney, D. J. (1996). Likability and similarity as potential sources of predictor-related criterion bias in validation research. Organizational Behavior and Human Decision Processes, 68(3), 272-286.

Schneider, B. (1987). The people make the place. Personnel Psychology, 40, 437-453.

Tornow, W. W. (1993). Perceptions or reality: Is multi-perspective measurement a means or an end? Human Resource Management, 32, 221-230.

Turban, D. B., & Jones, A. P. (1988). Supervisor-subordinate similarity: Types, effects, and mechanisms. Journal of Applied Psychology, 73, 228-234.

Wayne, S. J., & Ferris, G. R. (1990). Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: A laboratory experiment and field study. Journal of Applied Psychology, 75, 487-499.

Wexley, K. N., Alexander, R. A., Greenwalt, S. P., & Couch, M. A. (1980). Attitudinal congruence and similarity as related to interpersonal evaluations in manager-subordinate dyads. Academy of Management Journal, 23, 320-330.

Wexley, K. N., & Pulakos, E. D. (1983). The effects of perceptual congruence and sex on subordinates' performance appraisals of their managers. Academy of Management Journal, 26, 666-676.

Wexley, K. N., & Nemeroff, W. F. (1974). The effects of racial prejudice, race of applicant, and biographical similarity on interviewer evaluation of job applicants. Journal of Social and Behavioral Science, 20, 66-78.

Zalesny, M., & Kirsch, M. (1989). The effect of similarity on performance ratings and interrater agreement. Human Relations, 42, 81-96.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. American Psychologist, 35, 151-175.

# A Cross-Case Study: Revisiting Implementations of the Performance Analysis for Training (PAT) Model

*Catherine M. Sleezer Ph.D.*
*Oklahoma State University*

*Dale E. Kunneman Ph.D.*
*Meridian Technology Center*

*Five case studies are reported in the literature that describe systematic needs analysis using the Performance Analysis for Training (PAT) model. Using evidence from these reports, this cross-case study seeks more compelling and robust insights about needs analysis. The results of this study highlight differences and similarities in PAT model implementations. The results also provide insights into: (1) the correspondence between the PAT model and the actual needs analysis processes and results; (2) the influence of the organization, the decision-maker, and the analyst characteristics; (3) future practice and research needs.*

*Keywords: Needs Analysis, Performance*

The usual starting point for HRD and training efforts is the analysis of performance needs. The labels for this process include needs assessment, needs analysis, front-end analysis, and performance analysis (Rossett, 1992). The particular meanings of these terms have been debated (Moseley & Heaney, 1994), but each identifies a similar intent: To ascertain performance improvement problems or opportunities. Determining and then addressing performance needs is expected to produce benefits such as increased productivity, profit, satisfaction, and innovation. Given its importance, understanding and improving needs analysis is one strategy for improving HRD and organizational effectiveness.

## Problem Statement and Theoretical Framework

Practitioners have available many models that detail processes for analyzing needs (e.g., Goldstein, 1986; Gupta, 1996; Sleezer, 1991; Swanson, 1994). However, the literature contains few systematic studies that describe actual assessment practices in the workplace (Lewis & Bjorkquist, 1992; Tannenbaum & Yukl, 1992) even though many authors have indicated that such knowledge is needed. For example, Hatcher and Ward (1997) stated that because of the rapidity of organizational change, performance improvement professionals need innovative ways to analyze issues affecting performance. Also, Sleezer and Maile's (1997) study of needs analysis showed that HRD professionals used diverse practices and concluded by calling for additional research and model-building aimed at practical needs analysis tools.

A theory-based needs analysis model that has been tested in practice is the Performance Analysis for Training (PAT) model (Sleezer, 1990). Directed toward identifying the organization's performance needs that should be addressed with training, the model includes three components: (a) a conceptual framework with three elements that affect decisions about training needs, (b) a list of the phases and steps involved in the process of determining training needs, and (c) a set of worksheets that detail the phases, steps, and activities involved in determining training needs. The literature contains five case studies that each describe implementation of the PAT model in a functioning organization. Each study provides insight into one needs analysis effort and its context. Table 1 summarizes each case including the impetus for the needs analysis, the analysis level(s), the decision maker role and the analyst. Research-based case studies were implemented in four organizations: a financial services firm (Sleezer, 1991), a vehicle manufacturing division (Sleezer, 1996), a business forms manufacturing division (Kunneman & Sleezer, 1998) and veterinary medicine college (Knorr, 1997). The fifth case, which is a descriptive "how-to" case for practitioners, was implemented in a chemical manufacturing firm's strategic business unit (Sleezer, 1995). The levels of analysis varied across the five cases: one focused only on the organization level, two focused on the organization and the work-behavior levels, and two focused on the organization, work-behavior, and individual levels. Three analysts implemented the cases: Knorr, Kunneman, and Sleezer. Note in one case Sleezer led a team of analysts. As highlighted in Table 1, the combination of organization, decision-maker, and analyst characteristics is unique for each case. Each study served as a single case and resulted in an individual case report as specified by Yin's (1988) case study method.

*Copyright © 1999 Catherine M. Sleezer & Dale E. Kunneman*

**Research Questions**

This cross-case study reexamined the five case reports of PAT model implementations to answer the following research questions: (1) What is the correspondence between the PAT model and the actual needs analysis processes and results? (2) Whether the characteristics of the organization, the decision-makers, and the analyst influenced the analysis of needs? (3) What insight do the cases provide for future practice and research?

**Research Methodology**

This study relied on multiple-case design as described by Yin (1988). The methodology has three phases:

1. Design, which involves developing theory, designing data collection protocol, and selecting individual cases.
2. Single-case data collection and analysis, which involves conducting individual cases and writing individual case reports.
3. Cross-case analysis, which involves using the individual case reports to draw cross-case conclusions, modify theory, develop policy implications, and write the cross-case report.

This study focused on the final phase, cross-case analysis. To provide a context, we briefly overview the first two phases, which were completed prior to this study.

The design phase began with developing theory. As detailed in Sleezer (1990), the process included creating a needs analysis model that synthesized both theory and practical knowledge. It also included having three experts in training needs analysis evaluate the model's face and content validity. The outcome, the Performance Analysis for Training (PAT) Model, articulates needs analysis theory, operationally defines a needs analysis process and expected outcomes, and can serve as a logic model for case study research. A logic model stipulates a complex chain of events for an intervention over time and provides a basis for analyzing the data from a case study (Yin, 1998). Yin notes that the conceptualization "reflects the theory of the intervention" (p. 253). In the PAT model, the theory of the intervention is contained in the list of phases and steps involved in the process of determining training needs and in the set of worksheets. Conducting a needs analysis according to the list of phases and steps and the worksheets is expected to produce the following specific outcomes: (a) involve the organization's decision makers in the training process; (b) provide the organization's decision makers with accurate and sufficient information for making decisions about the contributions that training should make to the performance needs of the organization; and (c) provide information useful for the design, development, implementation, and evaluation of training. Rivals to this theory are that the phases, steps, and actions specified in the PAT model are impossible to follow, need to be modified, do not produce in the intended outcomes, or produce unintended consequences. A second theory is specified in the PAT model's conceptual framework: the analyst, the decision-maker and the organization characteristics interact to influence the needs analysis process and results. A rival to this theory is that the characteristics of one element dominate.

The design phase also involved creating a case study protocol and selecting individual cases. The PAT model Worksheets contain the protocol. The worksheets overviewed the project, and contained the procedures and general rules for implementing the PAT model in a functioning organization, and provided a template for organizing needs analysis notes. The case study sites and the analysts were selected based on feasibility and access. Individuals within each organization had negotiated with an analyst to conduct a needs analysis and also had indicated willingness for the project to serve as a case study. The analysts were knowledgeable about the topic. Sleezer had implemented many needs analyses, instructed needs analysis courses, and published articles on the topic. Kunneman and Knorr were doctoral students. Each had an association with the organization. Also, each analyst was willing to conduct the needs analysis using the PAT model and to participate in the research project.

The second phase of the research methodology was single-case data collection and analysis. This step involved conducting individual cases and writing the case reports. With this research method, each case that documented a holistic application of the PAT model in a context with distinct organization, decision-maker, and analyst characteristics tested the model and its theoretical base. Systematically studying needs analysis while implementing one required the analyst to engage in two on-going streams of activity. In one stream, the analyst implemented the needs analysis as prescribed by the PAT model. In the other stream, the analyst collected documentation and impressions about the process and results relative to the research questions and propositions.

The case study methodologies specified by Yin (1988) and by Miles and Huberman (1984) guided the data collection and analysis for the individual cases. While implementing the needs analysis, the analysts recorded notes

Table 1: Description of the Five Cases

| Case | Impetus for the Needs Analysis | Needs Analysis Levels | Decision Maker(s) | Analyst 's Name |
|---|---|---|---|---|
| Specialty Chemical Unit (SC): A strategic business unit of a chemical manufacturing firm. | Decision-makers wanted to identify future training needs for technical workers. The firm's strategic goal of quickly implementing new technology was the driving force for the needs analysis. | Organization Work Behavi Individual | Manager of Human Resources Plant manager | Sleezer |
| Veterinary Medicine College (VMC): A college in a US land grant university. | Decision makers responded to calls for action to all veterinary educational institutions from professionals writing in journals and speaking at conferences. They wanted to use a needs analysis as the first step in designing a curriculum that reflected customer needs. | Organization | Curriculum developmen t committee consisting of nine faculty members and the associate dean | Knorr |
| Business Forms Division (BF): A division of a worldwide manufacturer of business forms and business systems design. | Decision makers were implementing an ISO quality system designed to improve performance. The decision-makers requested the needs analysis as the first step in setting up an orientation program for new and current workers. | Organization Work Behavi | Plant Manager Chemical Dept. Manager | Kunnem n |
| Vehicle Manufacturing Division (VM): A division of a US manufacturer of vehicles. | The decision-maker was planning to implement integrated manufacturing throughout the division. He perceived that middle managers and supervisors knew that changes were eminent, but did not know how to manage in the new environment. He requested a needs analysis that would determine the critical needs for these employees. | Organization WorkBehavio | Director of Manufacturin g | Sleezer, project manager and tw project assistant |
| Financial Services Firm (FS): A US firm that sold financial processing products and services nationally. | The new training and development manager intuitively knew the training needs and thought that a formal training needs analysis might convince those in power to provide funding for training. | Organization Work Behavior Individual | President VP of Finance VP of Human Resources Training Manager | Sleezer |

relative to both streams of action. These notes provided the basis for the individual case reports. The cases emphasized different areas of interest based on the foci of the research questions and propositions. Four cases examined whether using the PAT model in the specific organization produce the predicted results. For three studies, implementing the PAT model was predicted to produce the results specified in the model. The VMC case examined the inclusion of the customer as an element in the PAT model and the usefulness of the project findings for making changes in the veterinary curriculum. Two cases focused on the correspondence between the PAT model and the actual analysis process, and two cases focused on whether the interaction among the model's elements influenced the needs analysis.

Cross-case analysis involves synthesizing the information from the individual cases, modifying the theory, developing policy implications, and writing the cross-case report. This study synthesized individual published case reports to examine multiple implementations of the PAT model in functioning organizations. Systematically examining evidence across cases indicates the extent of the replication logic. According to Eisenhardt (1989), Lee (1985), and Yin (1989), if cases are purposefully selected and multiple cases confirm predictions, then the propositions are supported. However, if a case contradicts the predictions in some way, then the initial predictions must be revisited and retested (Yin, 1989).

Synthesizing the information from the individual cases involved many steps. The first step occurred soon after the all the cases were complete with final reports written. The three analysts met and for the first time discussed the similarities and differences among the cases and the subtleties that they had encountered. Copious notes were taken during the meeting. Then to gain some distance, we allowed six months to elapse. The next step involved going through the files to get them in order. The data for the cross-case analysis consisted of the individual case reports and the notes from the analysts' meeting. After the data were ordered, we took long amounts of undisturbed time to review the information. The previously developed theories provided a template for comparing the empirical results of the case studies. We also developed coding categories based on the research questions and propositions, and we coded the data contained in the individual case reports. Then, we synthesized the coded data, searched for additional patterns to discover what could be learned to share with others, and compared conclusions from the individual cases to determine which information was replicated by other cases. Finally we used our notes to write this manuscript. Throughout the process, we checked and rechecked each other's work.

## Findings

We begin this section by examining the findings relative to the PAT model implementations and results. Then we examine the influence of the organization, decision maker, and the analyst characteristics on the needs analysis process and results. We conclude by identifying adaptations to the PAT model.

### *PAT Model Implementations and Results*

In all the cases, a decision maker within the organization determined the importance of a needs analysis project. A key event was the initial meeting between the decision maker(s) and the analyst. The project process and results were negotiated and agreement upon including such specifics as the use of the PAT model, data collection processes, time lines, resources, and reporting results. In some cases, the agreement included what would constitute a needs analysis. For example, in the BF and VM cases, the decision maker and the analyst agreed to begin the needs analysis project by implementing two PAT model phases and to table the third phase until after they obtained results from the first two phases. The VMC case revealed that the group of decision makers responsible for using information from the needs analysis to create a new curriculum was smaller than the group responsible for implementing the changed curriculum.

The iterative nature of the needs analysis process and the involvement of decision makers is apparent in the case reports. Four case reports (VMC, FS, SC, and VM) noted that the needs analysis process was iterative rather than strictly linear. For example, an external reviewer for the VMC case commented "There are multiple ways in which data are solicited. The assessment is an iterative rather than a cross sectional process, and this is good" (Knorr, 1997, p.73). Also, in all cases, decision makers participated in the process. For example, the VM Director commented that he had been involved in planning the analysis, had received in-process reports, and was "on top" of this project. He also appreciated not having to be the subject expert on performance analysis and having his time used wisely.

All three analysts used the worksheets as a guide and a framework for organizing information. For example, the SC team members individually filled in the worksheets and then met to compare notes. "This comparison produced many insights, as team members shared new information, revealed tacit assumptions, and clarified perceptions" (Sleezer, 1995, p. 247).

In each case, the analysts adapted the PAT model and used it to successfully implement a needs analysis project. One VMC adaptation was to add the customer to the PAT model elements. BF and VM adaptations included using information from the second phase of the PAT model to decide where to place the training emphasis. In both cases, the data gathered and synthesized about specific job behaviors resulted in transforming the decision makers sensed need into specific, delineated, and documented needs. The BF and VM cases also began Phase 2 before Phase 1 was complete. For example, the VM Director worked on clarifying perceptions from Phase 1, while the analyst began implementing Phase 2. Also, in the BF and VM cases, the need to implement Phase 3, individual capabilities analysis, was obviated by information that was gained during the first two phases. And in the BF case, the analyst expanded the PAT model's phase 2, Steps 2 and 3 by using a Valpar-2000 job profiling system rather than an analysis method that had been identified by Swanson and Gradous (1986).

The remaining adaptations focus on the PAT worksheets. The BF analyst commented on contextualizing the performance analysis tools and changing the order for actions. He noted that a question in Phase 1, Step 1, "Do the decision makers agree about the purpose and parameters of the analysis? was problematic because the parameters changed with the input of new data. The question was finally answered after Phase 2, Step 3. Similarly, the VMC analyst had insufficient information to complete the worksheets while the needs analysis was in process and so used an iterative approach that involved frequently reviewing previous responses to worksheet actions and expanding or recasting them when additional data or insight was received. He also contextualized the PAT worksheets, which were designed for use in business, to be appropriate for an educational situation. For example, he changed the terms 'training' and 'bottom-line' to 'curriculum revision' and 'goals'.

In all cases, the analyst and the decision maker (s) agreed that the expected results were produced. In three cases, BF, VM, and FS, the organization's decision makers explicitly stated that they were involved in the training process. Further, they received accurate and sufficient information for making decisions about the contributions that training should make to the performance needs of the organization, and information useful for the design, development, implementation, and evaluation of training. For example, the Vice-President of Finance at FS commented that he was ready to deliver the training in-house or find a vendor to deliver the content that had been determined. In the VMC case, the PAT model was modified to include the customer. The needs analysis project was expected to be useful for determining deficiencies in a professional curriculum, to involve the institution's decision makers in the process, and to result in changes to the curriculum. The analyst, the decision makers, and an external reviewer rated the project and determined that the results generally met expectations. One area of interest is the usefulness of the needs analysis findings for making changes in the curriculum. The decision makers rated the aspect as somewhat useful. However, the analyst observed the curriculum change process after the needs analysis project was complete and noted that the decision makers actually made few references to the information gathered from customers. He noted that while the modified PAT model was successful in guiding a needs assessment that produced the requested information; the decision makers made little use of the information; so the needs assessment findings contributed only slightly to changes in veterinary curriculum.

The analyst (s) and the decision maker (s) agreed that the implementation of the PAT model produced positive unexpected results. In three cases, BF, VM, and FS, the organization's decision makers reported that the needs analysis had changed their views of the role of training within their respective organizations. For example, the President at FS commented that he received information indicating a need for organizational restructuring to place training at a more strategic level within the organization. Another unanticipated result was the continued commitment from decision makers to implement the PAT model as a performance tool. The decision makers in the BF and SC cases commented on the usefulness of the analysts as future resources for performance analysis and training and requested the implementation of similar projects at other locations. Finally in all business cases, the decision makers concluded that the performance issues were more complex than they had initially thought and that providing training without considering other work place constraints would likely result in little use of the skills and knowledge.

The analysts in all cases concluded that the PAT model is most economical for guiding large-scale projects, such as reengineering or quality improvement, or for guiding affective projects, such as labor disputes. In such projects, awareness of diverse issues or feelings is critical to a success process and results, and the time and resources needed to implement the PAT model are appropriate given the high costs of poor performance. The PAT model may not be the best choice in less complex situations or when the performance results are relatively unimportant.

*Influence of organization, decision maker, and analyst characteristics on the process and results*

In each case the organization, the decision maker, and the analyst characteristics interacted to influence the needs analysis process and results. In the BF case, the analyst's experience with the Valpar-2000 job profiling system and his profiling certification influenced the method used to analyze jobs. In the VMC case, two organizational characteristics were influential: the veterinary medicine program had no established goal and the individuals who decided on and made curriculum changes held different perspectives about what the program should accomplish.

In all cases the interaction was as important as the element characteristics in producing effects. The BF case provides an example of how the context influenced the thinking of the analyst and decision maker, who in turn shaped the context: " the first phase, determining where the training emphasis should be placed, was influenced by the addition of new production operators to the organization. The analyst and the manager agreed that the training should be targeted at the production area because it had the most potential for individual and organizational performance" (Kunneman & Sleezer, 1998, p. 9). Interaction was also important in helping the individual stakeholders reach consensus. Such interactions occurred in all the cases. Two decision maker characteristics that interacted in important ways with the analyst and organization were the decision makers' support of the performance analysis and their power to implement both the training and non-training solutions influenced the project. In the VM, SC, BF, and FS cases, the interactions were positive. However, interactions were more problematic in the VMC case when individuals who were not members of the curriculum development committee had little influence on the needs analysis process, but significant influence on the curriculum revision.

*Adaptations to the PAT Model*

The case studies revealed aspects of the PAT model that worked well and adaptations to make it better reflect actual needs analysis processes and easier to use. Aspects that worked well include the model's components and its transparence. The analysts used the PAT model's conceptual framework, list of phase and steps, and worksheets to guide the needs analysis project. The worksheets provided direction for and a means of recording the needs analysis process and content. Further, they facilitated on-going reflection as the basis for insight. The PAT model components were virtually transparent to organizational decision makers who appreciated the wise use of their time. They were able to participate in the needs analysis projects without the burden of knowing all the details of the PAT model.

Adaptations to the PAT model include changing its name to Performance Analysis Tool, adding a "key population" element and the corresponding worksheets, and adding action planning as specific final step in the process. Another adaptation is to specify that a needs analysis does not necessarily include all three levels and that a work behavior analysis may be needed in order to determine where to emphasize performance. Yet another adaptation is to provide additional descriptive information on how to use the PAT model. The information could include the kinds of iterations to expect in implementing a needs analysis, strategies for gathering the information to complete each phase, strategies for adapting the sequence of actions in the PAT model to fit the context, use of the PAT model on large-scale or affective projects, and the probability of unexpected outcomes.

*Answering the Research Questions*

The cross-case results provide insight for answering the research questions and for modifying the theory and the PAT model. First, what insight do the findings provide on the correspondence between the PAT model and the actual needs analysis processes and results? In all cases, the analysts adapted the PAT model and used it to guide the analysis process. The adaptations varied across cases, but common themes emerged. One theme was that the process was more iterative and less sequential than indicated in the model. Another theme was that the worksheets were helpful in guiding the project and that they need to include additional information about strategies for using and adapting the PAT model. Yet a third theme was to include 'Key Population' as an element in the model. The PAT model implementations in all cases produced the expected results according to assessments by the analysts and decision makers; however, they also produced unintended positive results. Therefore, the theory of the intervention that is articulated in the PAT model should be modified to reflect the findings. Modifications should also include highlighting the iterative nature of needs analysis, providing additional information on using and

adapting the PAT model, adding 'Key Population' as an element, and describing the unintended results that can occur.

Second, what insight do the findings provide on whether the characteristics of the organization, the decision-makers, and the analyst influenced the analysis of needs? The findings confirm that the interaction of the organization, the decision maker and the analyst characteristics do influence the needs analysis process and results. The interaction was as important as the characteristics. No element dominated. The findings also identified two important characteristics: the decision maker's support for the needs analysis and power to implement solutions.

Third, what insight do the findings provide for future practice and research? Based on the cross-case findings, the PAT model should be modified so practitioners can use it to better guide their needs analysis projects. Future research can provide insights to continue honing the model. Future research could address the limitations of this study by using different research methodologies, researchers, and organizations. Areas that deserve research include determining when it is most feasible to use the PAT model. Because the PAT model requires extensive investigation, information is needed on when an organization benefits by investing in a PAT model needs analysis rather than a traditional, linear model needs analysis, or in taking no action at all. From a broader perspective, research on needs analysis effects from different perspectives is needed. The cases for this study provided information about the short-term results of needs analysis when assessed by the analysts and decision makers. What are the long-term results when assessed using concrete organizational performance measures or external evaluators? Future research could also examine how the capacity that is built while implementing a needs assessment is used. For example, consensus building among employees was an important aspect of needs analysis in the cases that were studied. How can those who are involved capture the process or shared insights to produce additional results?

## Conclusions

This study used evidence from five case study reports to obtain more compelling and robust insights about needs analysis. Systematic research on needs analysis can facilitate the continued improvement of needs analysis theory and practice with the ultimate goal of improving HRD and organizational effectiveness.

**References**

Eisenhardt, K. M. (1989). Building theories from castle study research. Academy of Management Review, 4(4), 532-550.

Gill, S. J. (1995). Shifting gears for high performance. Training & Development, 49(5), 25-31.

Goldstein, I. L. (1986). Training in organizations, needs assessment, development, and evaluation. Belmont, CA: Wadsworth, Inc.

Hatcher, T. G. & Ward, S. (1997). Framing: A method to improve performance analyses. Performance Improvement Quarterly 10,(3) 84-103.

Knorr, K. (1997). Adapting Performance Analysis for Training to Veterinary Medicine. Unpublished doctoral dissertation, Oklahoma State University, Stillwater.

Kunneman, D. E. & Sleezer, C. M. (1998). Using Performance Analysis for Training in an organization implementing ISO 9000 manufacturing practices: A case study. Academy of Human Resource Development Proceedings, USA, 301-308.

Lewis T. & Bjorkquist, D. C. (1992). Needs assessment - A critical reappraisal. Performance Improvement Quarterly, 5(4), 33-54.

Lee, A. (1985). The scientific basis for conducting case studies of organizations. Academy of Management Proceedings, 45, 320-324.

Miles, M. B. & Huberman, A. M. (1984). Qualitative data analysis: A sourcebook of new methods. Beverly Hills, CA: Sage.

Moseley, J. L & Heaney, M. J. (1994). Needs assessment across disciplines. Performance Improvement Quarterly 7(1), 60-79.

Rossett. A. (1992). Analysis of human performance problems In H. D. Stolovitch and E. J. Keeps (Eds), Handbook of human performance technology. San Francisco: Jossey Bass.

Sleezer, C. M. & Maile, C. A. (1997). Needs Assessment in Practice. Proceedings of the Academy of Human Resource Development. 290-296

Sleezer, C. M. (1990). The development and validation of the Performance Analysis for Training Model. (Project No. 39). St. Paul: University of Minnesota, Training and Development Research Center.

Sleezer, C. M. (1991). Developing and validating the Performance Analysis for Training Model. Human Resource Development Quarterly 2 (4), 355-372.

Sleezer C. M. (1995). Using performance analysis to assess future technical training needs. In Jack J. Phillips & E. Holton (Eds.), In Action: Conducting Needs Assessment. (pp. 241-253. Alexandria, VA: American Society for Training and Development.

Sleezer, C. M. (1996). Using performance analysis for training in an organization implementing integrated manufacturing: A case study. Performance Improvement Quarterly, 9(2), 25-41.

Swanson, R. A. (1994). Analysis for improving performance. San Francisco: Berrett-Koehler.

Swanson, R. A. & Gradous, D. (1986). Performance at work: A systematic program for analyzing work behavior. New York: Wiley.

Tannenbaum, S. I. & Yukl, G. (1992). Training and development in organizations. Annual Review of Psychology 43, 399-441.

Yin, R. K. (1989). Case Study Research: Design and Methods. Newbury Park, CA: Sage.

Yin, R. K. (1993). Applications of Case Study Research (Applied Social Research Methods Series Vol. 34). Newbury Park, CA: Sage.

# Large-Scale Training Needs Assessment of State Government Office Support Workers: A Case Study

*Donna H. Redmann*
*Louisiana State University*

*This paper reports on a case study of a large-scale, state-wide training needs assessment of office support/clerical employees in Louisiana state government. To identify the felt training needs, 11,016 surveys were distributed in a single mailing resulting a return of 7,117 completed questionnaires, a 65 % response rate. The paper□s focus is on the methodology, survey instrument development and administration, and how the resulting data was tabulated and analyzed to provided useful reports for HRD.*

*Keywords: Office, Training, Assessment*

Workplace demands for more competent and skilled employees have led to an increased need for training in both business/industry and state government. Numerous studies have been conducted that have assessed the training needs of, competencies required by, and tasks performed by office workers in business and industry. However, little has been done to examine the training needs of government office employees on a large scale (Alexander, 1996; Arney, 1998; Chalupa, 1997; Davis, 1991; Davis, 1992; Davis & Chaney, 1993; Marino, 1993; McEwen, 1997; Zhao, 1996).

Employees of the State of Louisiana are emerging from a long period during which little funding was available for employee training. A new state administration has recognized the strategic value of well-developed human resources. In 1997, the governor of Louisiana established work force development as one of the priorities for his administration. To achieve this objective, several groups were established: the Workforce Commission within the Governor□s office, a Task Force on Post-Secondary Technical and Adult Education, and the Occupational and Training Information System in the Louisiana State Department of Labor. The mission of the Occupational and Training Information Systems is to collect work force data for the Workforce Commission to use in setting priorities and funding decisions regarding training programs.

The first step required in any effective training intervention is a thorough and systematic assessment of workforce training needs. Successful needs assessments provide the basis for establishing the performance improvement value of subsequent training. Because of the highly varied nature of agencies and units in state government and the large volume of needs assessments that have to be conducted, Louisiana State University was contracted to assist in assessing and prioritizing training needs. During the initial funding period (the first six months), this needs assessment project consisted of two phases. Phase I was to identify the current training needs of approximately 12,000 incumbent office support/clerical employees in the state civil service system for all department and agencies throughout the entire state. Phase II was to develop a method for conducting needs assessment for all employees within a state department and to pilot this methodology with two departments (Holton, Bates, Redmann, Wilson, Naquin, & Burns 1997a,1997b, & 1997c.). The second period of funding (a two year period) is a continuation of Phase II which involves identifying the performance training needs of Louisiana state government employees for individual departments based on the methodology developed during the pilot study.

The focus of this paper will be on the large-scale training needs survey of state government office support/clerical workers that was conducted during Phase I of the initial funding period. The paper will cover the objectives of Phase I, the methodology employed to accommodate those objectives, including how the survey instrument was developed and administered, how the resulting data was tabulated and analyzed to provided useful reports for personnel managers. The paper will also explore the implication for Human Resources Development (HRD) training in government and business/industry. Only minimal emphasis will be placed on the actual findings of this large scale, state-wide survey.

## Phase I Objectives

The objective of Phase I of the Louisiana State Needs Assessment Project was to conduct a needs assessment to identify current training needs for office support/clerical employees within the state civil service system. The specific objectives were to identify the felt training needs of office support/clerical employees: (1) for all office

*Copyright © 1999 Donna H. Redmann*

support/clerical employees, (2) by job clusters, (3) for each state department, and (4) for each agency within the departments.

## Methodology

### *Instrumentation*

As a first step in the design and implementation of a training needs assessment program, it is essential to properly appreciate and understand the scope of the problem that was to be addressed. To this end, a preliminary series of meetings was held between the Phase I Project Coordinator and various State officials that were identified by the Director of the Comprehensive Public Training Program (CPTP), Division of Administration. These meetings were instrumental in exposing the complexity of job titles, job descriptions, GS levels, and other intricacies of the State Civil Service System. Furthermore, they helped form a foundation upon which the task of identifying office worker classifications that should be surveyed for possible training needs. The overall process required an examination of specifications for office/clerical job titles associated with series 4000 registers (no typing test required of applicant) and with series 4002 registers (typing test required of applicant). Ultimately, 45 job titles were selected for the training needs assessment survey of office support/clerical employees. Based on job descriptions and GS levels, these 45 job titles selected were subsequently grouped into the following eight clusters: Accounting, Administrative, Office Management, Clerical, Secretarial, Transcription, Word Processor, and University Admission.

The next step involved identification of the appropriate content of the survey instrument. Focus groups were employed to determine important areas for inclusion in the training assessment study; a total of eleven focus group meetings were conducted These focus groups consisted of state office/clerical workers representing multiple state agencies and encompassing a wide range of job titles. As a result of the 11 focus group meetings, it was discovered that there was considerable commonality of identified training needs across the groups. Therefore, it was decided that a single questionnaire should be developed. Also, it was anticipated that having one questionnaire would reduce the complexity of managing the distribution, collection, and tabulation of multiple versions of the instrument across the various state government departments and groups of office workers.

Based on the information obtained from the focus groups, a draft of the questionnaire was developed. A panel of experts was employed to review the questionnaire for content and design. A field testing of the survey instrument was conducted with 16 office/clerical workers from the Department of Environmental Quality (DEQ), representing a wide range of job titles and various agencies within DEQ. The volunteers were asked to provide any suggestions or recommendations that would improve the instruments, to check the wording of the items, to make sure the directions were clear and understandable, and to keep track of their time in order to estimate how long it would take the typical respondent to complete the questionnaire. The average time to complete the survey instrument was 15 to 20 minutes.

The final draft of the survey instrument contained a total of 67 items covering nine training topic areas, 3 questions dealing with computer and related technology, space for writing open-ended recommendations and comments, and a section on demographics. The nine topic areas consisted of oral communication, written communication, files/filing, general skills, human skills, personnel issues, understanding general state procedures, understanding one□s own department/agency, and work management skills. For each of the 67 training topics listed, the respondents were asked to use the following scale to indicate how much their job performance would improve if they were to be provided training in the topic listed: 1 = None, 2 = A Little, 3 = Somewhat, 4 = A Lot, 5 = A Great Deal, and N = Not Applicable. The frequency response to this scale was employed in the ensuing statistical analysis to provide a ranking of the relative importance of the various training topics addressed.

In addition to the □pure□ training topics covered above, the survey instrument included a section on computer and related technology wherein respondents were asked to identify the version of listed software used in performing their job. Further, they were asked to indicate their training needs for basic computer literacy, mainframe, a wide range of application software (e.g. WordPerfect, Lotus, PowerPoint), electronic mail, and the Internet by using the following scale: 1 = Don□t need training, 2 = Completed basic training which was adequate, 3 = Need basic training, and 4 = Need more advanced training. The respondents were also asked if they would be interested in being trained to be an in-house resource person to help co-workers with computer software/hardware problems. Finally, in

the demographic section of the questionnaire, the respondents were asked to identify their job title from among the list of 45 job titles surveyed. Additional questions dealt with work location, number of years spent working for the Louisiana state government, the number of years in current position, and level of education.

To reduce overall respondent error rate and encoding errors, a Scantron sheet format was employed for the four-page survey instrument. The respondents were instructed to use a number two pencil to bubble in their responses. A side benefit of using a Scantron sheet format was that the respondents□ time to complete the questionnaire would be lessened somewhat, therefore, it was anticipated that they would spend less time away from their job and would be more likely to complete and return the survey.

To ensure anonymity, respondents were asked to identify their department, agency, and work location only. Upon receipt of the completed questionnaire, the CPTP staff assigned the appropriate organization I.D. number to a questionnaire I.D. code sheet. This Scantron I.D. code sheet was designed to accompany the questionnaire for the purpose of associating the individual questionnaires with the appropriate organizations.

*Administration of the Survey Instrument*

The staff at the CPTP distributed the questionnaires to contact persons (human resource director or related person) at 272 state agencies. These contact people then distributed the questionnaires to all their office support/clerical workers. The contact people were asked to return the completed questionnaires within 3 weeks. Due to a time limitation, a questionnaire follow-up was not conducted. Of the 11,016 questionnaires distributed statewide, a total of 7,117 completed questionnaires were received. The overall response rate for the State was 65%.

**Analysis, Interpretation, and Format of Survey Results**

For the eight job clusters representing the 45 job titles, statistical analysis software (the SAS System for Windows, Release 6.11) was employed to: (1) tabulate the scale response frequency for each questionnaire training topic; (2) rank order the 67 questionnaire training topics, based on an assigned score which is the percentage of the respondents that rated that item in either of the two categories of highest importance to job performance (□ □Lot and A Great Deal); (3) generate frequency and percentages responses for basic and advanced computer training needs, for training as an in-house resource person, and for demographic information.

To recommend training for the 67 topics, the following criteria were used: If the ranked order score received was 60 % or higher, then there was a perceived Widespread Need for training related to that item; if the score was between 40% and 59.9%, then there was Selected Group Need; and if the score was below 40%, then there was Individual Employee Need for that training item. For example, a score of 66.3 indicates that 66.3 % of the respondents rated the item either as of A Great Deal or A Lot with regard to its importance relative to their job performance. Therefore, this training item would be identified as a widespread need for training.

For the basic and advanced computer training needs recommendations, the need basic training and the need more advance training response categories were reported separately in rank order with the highest need listed first. The demographic information for the global responses (all 7,117 responses) was reported by the eight job clusters and for the 272 individual agencies, the 45 job titles were used. For example, 24 job titles were reported for the Department of Revenue and Taxation; 17 for the Department of Public Safety Services.

Due to the comprehensiveness and large scale nature of this state-wide training needs assessment of office support/clerical employees, the survey findings were delivered to CPTP in the following formats:

1. Global responses, based on all 7,117 completed questionnaires received
2. Per job cluster, based on the eight different groups of job titles
3. For each of the 272 State organizations as identified by organizational I.D. number.

The information provided in the first two formats were for use by CPTP in planning statewide and/or multi-agencies training. The third format was provided to each of the 272 agencies/organizations with a guide/key for interpreting the results.

During the initial funding period (the first six months) two organizations, Department of Public Safety Services and Department of Revenue and Taxation, participated in Phase II where an in-depth performance-driven training needs assessment methodology was piloted (Holton et. al., 1997a & 1997b). The pilot study resulted in a multi-level, multi-dimensional strategic approach that employed both a top-down and a bottom-up strategy. The individual agency reports of the results from the office support/clerical employees survey from Phase I was one of the multiple data collection methods used in the Phase II agency in-depth training needs assessment pilot study.

**Findings and Procedural Recommendations**

Due to the sheer magnitude of this large scale needs assessment, the findings based on the above analysis will be highlighted in terms of the methodology.

***Return Rate***

A 65 % return rate (7,117 out of 11,016) is considered to be successful for a one-time mailing of the questionnaire with a 3 week turn-around. Several factors may have contributed to the high response rate. First, the 11 focus groups involved 93 office support employees representing the different job clusters and different governmental agencies. Since the members were asked to provide input in the development of the survey instrument, there was tremendous buy-in. These same employees may have helped to motivate their fellow workers in completing the questionnaire. Second, the instrument was designed to be read and completed easily and efficiently. By having the respondents bubble in their responses on the Scantron sheet format instrument, the actual time to complete the questionnaire was reduced. Third, the rationale section of the instruments was written to motivate the employees to complete the questionnaire by stating that they were being given the opportunity to express their perception of what type of training would help them in improving their job performance. Fourth, insuring anonymity may have also contributed to the return rate. Fifth, advance notice of the survey was sent to those individuals responsible for distributing the questionnaire at their agency.

Examining the return rate for the different job clusters, the largest number of global responses (all 7,117 respondents) were received from the largest cluster of office workers, the clerical group (2,700 questionnaires received). The highest global response rate by cluster was office management, 70.3 % (433 out 616) (See Table 1.). In comparison to the two pilot agencies, the Department of Revenue and Taxation and the Department of Public Safety, the highest number of responses were from job titles representing the administrative cluster. Variation from agency to agency is expected due to the individual mission of each organization.

Table 1
Questionnaire Response Rate by Job Clusters

| Job Cluster | No. Sent | No. Received | Response Rate | No. of Job Titles |
|---|---|---|---|---|
| Accounting | 326 | 182 | 55.8 % | 1 |
| Administrative | 1,016 | 655 | 64.5 % | 10 |
| Clerical | 4,274 | 2,700 | 63.2 % | 8 |
| Office Management | 616 | 433 | 70.3 % | 6 |
| Secretarial | 3,217 | 2,024 | 62.9 % | 8 |
| University Admission | 99 | 46 | 46.5 % | 2 |
| Transcription | 86 | 4 | 4.6 % | 1 |
| Word Processing | 1,382 | 912 | 65.9 % | 9 |
| Missing Titles | 161 | | | |
| Total | 11,016 | 7,117 | 64.6 % | 45 |

*Training Topics*

Based on the ranking of assigned score for the 67 questionnaire items dealing with training topics from all 7,117 responses received, the respondents indicated that there were no topics that would have widespread need (a rank order score of 60 % or higher), but did indicate 20 training topics that had selected group need (a rank order score between 40 % and 59.9 %), and 47 training topics that had individual employee need ( a rank order score below 40%). The 20 selected group training needs represents the following 7 topic areas: written communication, files/filing, general skills, human skills, personnel issues, understanding your department/agency, and work management skills. The 47 identified individual employee training needs represent all 9 training topic areas.

If statewide, multi-agency, and or individual agency training is to be offered to office support staff in similar job titles, then providers of performance improvement training activities may want to examine the survey findings per job clusters. The findings, from the 7,117 completed questionnaires, indicated that widespread training needs were identified for 5 of the 8 job clusters and these needs were grouped into the following 3 topic areas: human skills, personnel issues, and files/filing. Table 2 list the widespread and selected group training needs by topic areas for all respondents and for each job cluster. The human skills topic area was rated most frequently as a widespread training need. Stress management training, one of the 10 human skills training topics, was ranked first for 4 job clusters (accounting, clerical, university admission, and transcription) and fourth for office management. Oral communication appears to be the least needed training topic for the selected group needs. The identification of training topics by job cluster, as presented here, can also be useful to curriculum planners at vocational business and office training institutions.

Table 2

Widespread (W) and Selected (S) Group Training Needs Identified by All Respondents and by Job Clusters

| Topic Area | All | Acc□t | Admin | Clerical | Office Mgmt | Secr | Univ Admis | Tran-script | W.P |
|---|---|---|---|---|---|---|---|---|---|
| Oral communication | | | | | S | | | S | |
| Written communication | S | S | S | S | S | S | | W | S |
| Files/filing | S | S | S | S | S | S | S | S | S |
| General skills | S | S | S | S | S | S | S | S | S |
| Human skills | S | W-S | S | W-S | W-S | S | W-S | W-S | S |
| Personnel issues | S | S | S | S | W-S | S | S | W-S | S |
| Understand general state procedures | | | S | | S | | S | S | |
| Understanding one□s own agency | S | S | | S | S | | | S | S |
| Work mgmt skills | S | S | S | S | S | S | S | S | |

At the agency/organizational level, the recommendations for training may differ from the global findings as illustrated in Table 3. The table lists the top 5 items for all respondents, for the Department of Public Safety, and for the Department of Revenue and Taxation.The respondents from the Department of Revenue and Taxation indicated that there were 5 training topics that would have widespread need, 31 training topics that have selected group need, and 31 training topics that have individual employee need. For the Department of Public Safety, 5 training topics were identified as having widespread need, 17 topics as having selected group needs, and 45 as having individual employee need. The two training topics where there appears to be agreement on the rank order (within the top five rankings) are related to human skills: stress management and dealing with difficult people who are co-workers. The remaining top-ranked topics are related to personnel issues, understanding your department/agency, and work management skills. At the other end of the scale, there was agreement that oral communications skills training was required only at the individual employee need level. Overall, these findings would tend to indicate that there is not a strong need to provide the same training to all agencies throughout the State. Therefore, the agency level of rank ordering of employees felt needs for training should be given a higher priority than the global ratings when identifying the appropriate performance improvement intervention at the department/agency level.

Table 3

Rank Order of Selected Training Topics by Office Support Workers

| Training Topic | Global | Public Safety & | Revenue Taxation |
|---|---|---|---|
| Stress management/stress relief. | 1 | 2 | 3 |
| General knowledge of Civil Service Rules. | 2 | 34 | 1 |
| Dealing with difficult people who are co-workers | 3 | 3 | 5 |
| Computerized filing systems | 4 | 12 | 25 |
| Dealing with difficult members of the public | 5 | 34 | 21 |
| Cross training to fill in for other employees in your department. | 13 | 34 | 2 |
| How to locate and retrieve information | 16 | 1 | 8 |
| How to handle multiple tasks. | 18 | 4 | 4 |
| Decision making skills | 14 | 5 | 6 |

***Software Usage***

With regard to the type and version of software being used by the office support employees in their jobs, the two pilot departments appeared to agree in general with the global responses. When asked what type of computer training they need in order to perform their jobs more effectively, they indicated a need for basic training regarding the departmental mainframe and in some of the newer technology, such as the Internet and e-mail. All three groups of employees agreed on their perceived need for more advance-computer training on one application software, WordPerfect. The remaining 15 advanced computer training needs varied widely; therefore, computer training needs should be identified at the department/agency level. It is interesting to note that the survey revealed that many employees were still using DOS and original versions of Windows. Since these technologies are changing so rapidly and the fact that State government is committed to upgrading technologically, it is recommended that computer training needs be assessed more frequently.

***In-House Trainers***

To determine whether any of the office support workers would like to be trained as an in-house resource person helping co-workers with computer software/hardware problems, a question to this effect was included in the survey instrument. Almost forty percent (39.6%) of the 7,117 respondents indicated they would like to receive this type of training. Over a third (38.6 %) of the respondents were not interested and 21.8 % responded in the negative because there was someone already trained in their department to serve in that capacity At the departmental level, over half

(50.5 %) of the Department of Public Safety and 28 % of the Department of Revenue and Taxation respondents expressed an interest becoming an in-house resource person.

**Implications**

The approach utilized and the unique problems encountered in this case study of an unusually large- scale need assessment of office workers may be useful to other organizations in their move to performance driven human resource development. Examples of the unique problems associated with this state government needs assessment were:

- ALL employees throughout the entire state were to be given the opportunity for input.
- 45 different job titles and civil service job descriptions were to be included.
- The reading levels of the office support staff had to be considered.
- Political implications, e.g., all state government departments though out the entire state had to be included in the study.
- This state-wide needs assessment had to be conducted within a six month period.

Effective performance related need-assessments should to be linked to multi-dimensional strategies that include both top-down and bottom-up approaches to collecting information. The wide-scale survey used in this case study did offer an efficient means for surveying an extremely large number of employees in a short period of time. According to Holton (1995), large-scale □felt needs□ training assessment can be effective at boosting morale, but may have minimal success in improving performance. Therefore, the results from a large-scale felt needs survey should not be the only source used for identifying performance improvement training.

**References**


Alexander, M. W. (1996). The secretary□s new role as trainer: Implications for office educators. Office Systems Research Journal., 14(2), 19-25.

Arney, Janna B. (1998). A comparison of computer skills needed by business administration graduates as perceived by colleges and universities and human resource managers in Indiana. NABTE Review, 25, 44-48.

Chalupa, M. R. (1997). The status of presentation software usage in the workplace. Office Systems Research Journal, 15(1), 33-41.

Davis, D. C. (1991). Status of in-house training programs for office support staff with implications for office systems curricula. Office Systems Research Journal, 10(1), 15-21.

Davis, D. C. (1992). Skills needed by office support staff as perceived by trainers. The Delta Pi Epsilon Journal, 34(1), 13-23.

Davis, B. D., & Chaney, L. H. (1993). Problems of corporate training and development personnel. Office Systems Research Journal, 11(3), 26-33.

Holton, E. F. III (1995). A snapshot of needs assessment. In J. Phillips & E. Holton (Eds.), In action: Conducting needs assessment (pp. 1-2). Alexandria, VA: ASTD Press.

Holton, E. F., Bates, R. A., Redmann, D. H., Wilson, J. A., Naquin, S. A., & Burns, N. J. (1997a). LA state needs assessment project: Department of public safety services training needs. (Technical Report, Vol. II). Baton Rouge, LA: Louisiana State University.

Holton, E. F., Bates, R. A., Redmann, D. H., Wilson, J. A., Naquin, S. A., & Burns, N. J. (1997b). LA state needs assessment project: Department of revenue & taxation training needs. (Technical Report, Vol. III). Baton Rouge, LA: Louisiana State University.

Holton, E. F., Bates, R. A., Redmann, D. H., Wilson, J. A., Naquin, S. A., & Burns, N. J. (1997c). LA state needs assessment project: Office support/clerical training needs. (Technical Report, Vol. VI). Baton Rouge, LA: Louisiana State University.

Marino, P. (1993). Job design and skill level in using information technology: Perceptions of administrative support personnel in financial organizations. The Delta Pi Epsilon Journal, 35(3), 121-137.

McEwen, B. C. (1997). Preparing office professionals for the next century. Business Education Forum, 52(1), 42-45.

Zhao, J. J. (1996). Recommended computer-end-user skills for business students by Fortune 500 human resource executives. The Delta Pi Epsilon Journal, 38(3), 155-166.

# Results Assessment System: From Evaluation to Assessment

*Richard A. Swanson*
*University of Minnesota*

*Elwood F. Holton III*
*Louisiana State University*

*Critics and scholars note that the HRD profession is stuck on a 4-level evaluation model that still doesn't work after 30 years. Practitioners know about the importance of evaluation and, yet, are not evaluating the results of their efforts. As an alternative, the Results Assessment System is firmly grounded in Human Resource Development theory and research. This paper describes the system and critiques it from the perspectives of credibility and practicality.*

*Keywords: Evaluation, Assessment, Results*

Human Resource Development (HRD) professionals achieve results in organizations by developing or unleashing expertise in individuals for the purpose of improving performance-- at the organizational, work process, work group, and/or individual levels. The *Results Assessment System* allows researchers and practitioners assess those results (Swanson & Holton, 1999). More specifically, it directs the assessment of performance, learning, and perception results from HRD interventions and other development efforts sponsored by leaders, managers, and workgroups.

## Assessment Problem

Critics and scholars note that the HRD profession is stuck on a 4-level evaluation model that still doesn't work after 30 years. Practitioners know about the importance of evaluation and, yet, are not assessing the results of their efforts (Swanson & Holton, 1999)

HRD has not been a serious performance partner in most organizations, and thirty years of the 4-level evaluation model has done very little to change this condition. This condition leaves top managers and clients typically asking for nothing more than participant reaction evaluations-- 'smile sheets' that do not correlate with either learning or performance. When top managers finally decide to do a serious assessment, it is usually in response to a crisis or to their own judgement that "things have gone awry," and invariably comes too late.

## Background

Go to any bookshelf of management or organizational books and look up the word "evaluation" in the index. How often do you think the word is mentioned? We can't find it. Yet, members of the HRD profession consistently talk about evaluation while rarely practicing it in the field. What is going on here? One clue is in the terms found in the indexes of most business books-- words like performance, assessment, and measurement. In contrast, evaluation professionals tend to be consultants who study educational and government programs rather than focus on core organizational performance and processes. We have come to realize that the word "evaluation" evokes a negative schoolhouse picture and the idea of assigning grades to ill-defined activities.

In contrast, the *Results Assessment System* is about assessing results, which is different than evaluation. *Assessment* of results is a core organizational process. In contrast, surveys of evaluation practices in HRD clearly show that HRD practitioners view evaluation as optional (Bassi, et. al., 1996). The well-known 4-level evaluation model (Kirkpatrick, 1994) has failed the profession for a number of reasons (Alliger, et. al., 1989 & 1997; Bassi, et.al.1996a & 1996b; Dixon, 1990; Holton, 1996; Newstrom, 1996; Swanson, 1997). A key failure is its emphasis on "reactions" versus the fundamental performance results of the host organization.

Readers will feel initial comfort with the *Results Assessment System* relative to the familiar *4-Level Evaluation Model*. At first glance they are not radically different from one another. A second glance reveals fundamental, substantial differences in (1) their evaluation and assessment priorities and (2) the completeness of the practical research-based process and tools offered in the *Results Assessment System*.

*Copyright © 1999 Richard A. Swanson and Elwood F. Holton III*

**Purpose of the Paper**

The purpose of this paper is to overview the *Results Assessment System* and to critique it in terms of its credibility and practicality in assessing results from human resource development efforts. The *Results Assessment System* claims that it can be applied to any HRD program in any organization.

**Overview of the Results Assessment System**

One of the reported strengths of the *Results Assessment System* is that it is an integrated system that is firmly grounded in Human Resource Development (HRD) theory and research. The purpose of this section is to provide an overview of the system, and frame it in the larger HRD development and improvement process. These core concepts provide the foundation for each chapter that follows.

**Human Resource Development Process**

The process of change and improvement within organizations, work processes, and individuals is fairly universal. While each specialization may have its own vocabulary and variations, the core process does not change much. Human resource development can be characterized as having the five phases (Figure 1).

Almost all of the models of HRD or organizational performance change contain a final phase in which outcomes are assessed. Unfortunately, the final phase is the one most often ignored. It is the final phase upon which this paper focuses.

**Figure 1: The Human Resource Development Process**

| Phase | Description |
|---|---|
| 1-Analyze | *Determine performance requirements and their relationships to HRD interventions.* |
| 2-Propose | *Design strategies and propose appropriate interventions.* |
| 3-Create | *Create intervention components-e.g., design, logistics, program, materials, etc.* |
| 4-Implement | *Implement the intervention-e.g., organization development and/or training.* |
| 5-Assess | *Determine if desired results have been achieved.* |

It is also critical to recognize that HRD professionals work their HRD process in context of a host organization. This is illustrated through a 'world view' model of HRD in which the five-phase HRD process is laid in the organization and parallels other organizational processes.

This systems world-view model has value for clarifying the important role of results assessment in HRD. HRD is a means to achieve some larger result in an organization. It is for this reason that results assessment is not optional, for it is the results that impact the organization, and ultimately provide outputs desired by customers. For HRD to be a core organizational process, it then follows that results must be assessed. This model also has utility in keeping the big issues surrounding results assessment in perspective in that the world does not begin and end with HRD.

Familiar models of improvement, change, and development are variations of this core development model. For example, Lewin's organization development model (unfreeze, move and refreeze) is an elementary view of the journey of change with assessment driving the process. It is fun to think of this and related HRD processes from both the individual perspective and the total organizational perspective of moving from and existing place to a new place. Almost all of the HRD models are variations of the core five-phase HRD process and call for assessment of results.

*Results Assessment System Components*

Results assessment reflects the last phase of the five-phase HRD process. Results assessment can be defined *as determining if desired outputs have been achieved by the HRD interventions*. This book presents a system for assessing results from HRD interventions, called the *Results Assessment System*. The system has four basic components: a *process*, result *domains*, a *plan*, and *tools*. As shown in Figure 2, the process describes the general steps necessary to conduct a results assessment in an organization. It is described in more detail below. Domains define the possible range of results that can be assessed. The plan identifies all the key decisions that must be made to design a results assessment. The tools provide the means to actually measure results.

| **Figure 2 - Components of the Results Assessment System** | |
|---|---|
| PROCESS | What are the general steps necessary to conduct a results assessment in an organization? |
| DOMAINS | What domains of outcomes should be assessed within HRD? |
| PLAN | What are the key decisions that must be made to design a results assessment, including outcome domains to assess? |
| TOOLS | How are results measured? |

**Results Assessment Process**

The core process of the *Results Assessment System* is illustrated in Figure 3. This five-step process view of assessing results provides a simple and direct journey from analysis inputs to decision outputs.

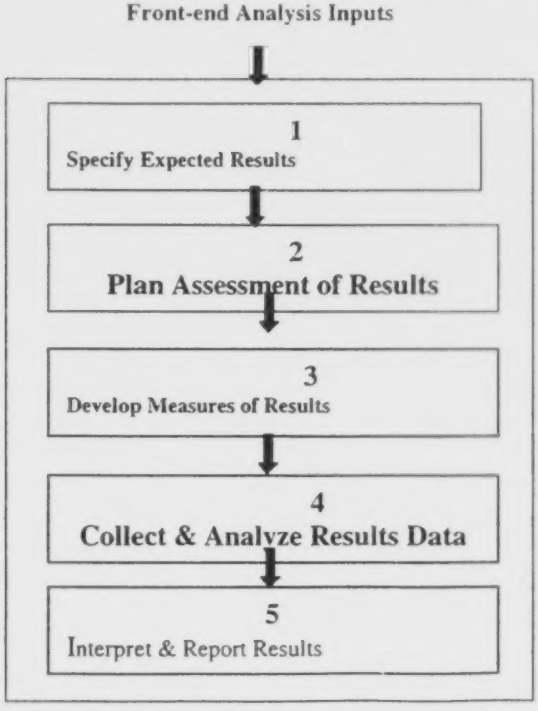


**Figure 3**

Results Assessment Process

**Decision Outputs**

Several aspects of this process are worth noting here:

1. To be most effective, results assessment requires strong front-end analysis. It is hard to assess outcomes if the goals are not well defined.
2. This process has an action orientation in that the final outcome is presumed to be decisions about HRD interventions and actions taken to maintain, change or eliminate programs. The purpose is not research.
3. The five steps present a simple and logical progression of activities essential to effective results assessment.

Considerable emphasis is placed on defining expected results and properly interpreting them, in addition to the issues of developing good measures.

*Results Domains*

The *Results Assessment System* is capable of measuring results within three domains: *performance*, *learning*, and *perceptions*. Each domain has two options within it. These domains and options within *the Results Assessment System* are specified below.

**Performance Results** are defined as:

**System:** The units of mission-related outputs in the form of goods or services having value to the customer and that are related to the core organizational, work processes, group/individual contributors in the organization.

**Financial:** the conversion of the output units of goods or services attributable to the intervention into *money and financial interpretation.*

The assessment of system, mission-related outputs are bottom-line talk. Things like 50,000 cars made; the patient lived, or 5,000 service contracts sold. Those pursuing improvements in an organization face a key dilemma--mission-related outputs. While the focal point is the output goal, the means or driver of that goal may look very different. A simple analogy would be a sporting event in which the goal is to win by getting the ball through the hoop more times than the opponent-- scoring points.

While scoring points and enough to beat the opponent is the mission-related output, there are a wide variety of development actions-- *performance drivers*-- that could be pursued by the coach, assistant coaches, the team owner, the offensive players or strategy, the defensive players or strategy, and/or individual players. Connecting the performance drivers to the system outputs of performance is critical to truly understand performance results.

Without front-end analysis, decision-makers in complex organizations will often implement simple-minded and/or invalid performance drivers that have no real connection to the system performance goals. The book "Witch Doctor's: Making Sense of Management Gurus," by Micklewait and Woolridge (1996), is an extended analysis of the false promises made by well meaning people about their "special brews" *(performance drivers)* without any substantial evidence of their connection to the system outputs *(performance outcomes)*.

**Learning Results** are defined as:

**Knowledge:** Mental achievement acquired through study and experience.

**Expertise:** Human behaviors, having effective results and optimal efficiency, acquired through study and experience within a specialized domain.

Knowledge, an intellectual or cognitive result of learning experiences, is the basic learning result. It is lodged in a person's mind. Measures of knowledge confirm the level of knowledge held by individuals within a particular subject area. In terms of being both effective and efficient, paper and pencil tests are the primary means of measuring knowledge.

Human expertise is the second category of learning -- the more complex learning result category. People with expertise have knowledge and are able to act upon that knowledge. The effective and efficient ability to act upon knowledge generally comes from experience beyond core knowledge. Measuring human expertise requires that an individual demonstrate their behavior in a real setting or a simulated setting. When assessing learning

results, we generally recommend that both knowledge results and expertise results be measured. And, it is logical that knowledge can be measured some time before expertise in that the learner needs time to gain experience. The span of time will vary depending on the complexity of the area of expertise being developed.

**Perception Results** are defined as:

**Participant Perceptions**: Perceptions of people with first-hand experience with systems, processes, goods or services.

**Stakeholder Perceptions:** Perceptions of leaders of systems or people with a vested interest in the desired results and the means of achieving them.

Of the three domains, perceptions have the lowest cost and lowest return. It is lowest in terms of cost in the sense that simple, short, and standardized perceptions rating forms can be produced for participants and stakeholders. It is lowest in terms of return because it provides the least valid information about performance outcomes.

Perception results are perceptual states held by human beings that you are concerned about. Measures of perceptions systematically access this information from selected groups of people. The mantra for perception results domain should be (1) get the data, (2) do not spend a disproportionate amount of resources to get it, and (3) do not over-interpret it. It is recommended that you get this data only as long as it is not used as a substitute for measuring learning results and/or performance results. For example, people self-reporting that they have learned something is not a measure of what they learned. It is also recommend that perception results data from both the participants and the stakeholders be obtained.

## Critique of the Results Assessment System

Many times, results assessments that are practical turn out not to be credible (and vice-versa). The challenge is to plan an assessment that is *both* credible and practical. This section discusses the credibility and practicality of the *Results Assessment System* from the perspectives of researcher and practitioner.

Credibility can be thought of as the power to elicit belief. For HRD the *belief* in the Results Assessment System process and outcomes must be elicited from both assessment experts and organizational practitioners. Practicality can be thought of the ability to be used. For HRD the *usefulness* of the Results Assessment System process and outcomes must be elicited from both assessment experts and organizational practitioners. The Results Assessment System purposefully addresses the difficult issues of gaining credibility and practicality approval from both camps.

### *Framing the System*

Big-picture credibility of the Results Assessment System is framed by a four component assessment system (see figure 2) and the positioning of assessment as a 5-step process (see figure 3) that details the last phase of the 5-phase human resource development process. While the presentation of the conceptual integrity of this framework is lean and will likely be criticized by assessment experts, we estimate that it will withstand their criticism. It is important to note that their intellectual reach is from in-depth assessment expertise. With little knowledge of assessment, the critical lens of the practitioners will most likely concern itself with *face validity* and *simplicity*. Practitioners that do not view their work from a process orientation will either have a new realization about the systems nature of their work or experience consternation when they consider the *Results Assessment System*. The popular options for practitioner's are Phillip's overwhelming 18-phase model (1983) and Kirkpatrick's shallow 4-level model (1998).

### *Consistency and Accuracy*

Measures have to possess two key characteristics to be of value: they must be *accurate* and *consistent*. Measurement experts call these two concepts validity and reliability. Measures are said to be accurate if they measure what they are supposed to measure. For example, self-reported measures of performance on the job tend to not be very accurate because people tend to overrate themselves. A consistent measure is one that yields the same results every time it is used. A measure can be very consistent but not valid (measure inaccurately or the wrong thing). For example: a person may be quite consistent in his or her performance rating, but not very accurate.

Measures used to assess performance, learning, and perceptual results need to be both reliable and valid to be of any value. Both reliability and validity are dealt with in the Results Assessment System. In contrast, Kirkpatrick's latest book does not deal with either issue (1998). Neither reliability nor validity are even in his table of contents let alone dealt with dealt with in the text. This is one of the several fatal flaws in the Kirkpatrick model of evaluation. Evaluation practices that do not consider reliability and validity are irresponsible and reckless.

Measurement experts talk about reliability as a pre-condition to actual validity. There are general rules for building in reliability and statistical means of checking it out. The *Results Assessment System* informs readers how to build in reliability and provides minimal guidance as to how to statistically calculate reliability. The guidance in building in reliability should satisfy the critical eyes of both assessment experts and practitioners. As for the statistics, the system defers to the use of standard data analysis programs and this slim presentation will not likely satisfy assessment experts.

Measurement experts talk about three common types of accuracy (or validity). The minimum and most important requirement is that the content of your measure matches the content of what you are trying to measure (content validity). For example, for a performance-rating instrument, the items on the instrument must minimally match what is really required to do the job. This is usually established by subject matter experts and is done logically, not statistically. The *Results Assessment System* provides simple and direct guidance in this realm to gain support from both assessment experts and practitioners.

A higher-level accuracy check is to ask whether the measure really predicts the dependent variable that it is supposed to predict (criterion validity). Thus, we would expect a performance-rating instrument to be able to predict, or distinguish, high performers from low performers. If we find that high performers in an organization have widely varying scores on the performance-rating instrument, then it would not have good criterion validity. An instrument could have accurate content (appear to have the right content) but not be a good predictor, probably because important things were left off the instrument. Again, the *Results Assessment System* informs readers how to establish such validity and provides minimal guidance as to how to statistically determine it. The system them again defers to the use of standard data analysis programs that will not fully satisfy readers who are assessment experts.

***Planning Practical Assessments***

It is actually easy to plan an assessment that is expensive, difficult to use, and is ultimately not executed and/or utilized. There are so many rival ideas about evaluation and measurement that it can very quickly become confusing. In fact, the primary reason for the six-option results menu of the *Results Assessment System* is to reduce the overwhelming number of options that experts in measurement and evaluation put before us. This framing of the options to six makes assessment work manageable from the very beginning. Creating an endless assessment options list causes confusion that almost immediately threatens the principles of practicality and credibility.

Many professional evaluators forget that the job is not evaluation, but that job of measurement and assessment to be help in getting something more fundamental done—producing results. Having too many options partially explains why many newcomers to measurement say that "it costs too much" or "it takes too much time" and why simple, inadequate schemes are attractive and usually selected by them. The 1-page *Results Assessment Plan* of the *Results Assessment System* chapter guides the basic decisions that will help researchers and practitioners build a measurement system to meet their desired ends. This critical tool, as much as any in the *Results Assessment System*, is pivotal in getting the in reining in both scholars and practitioners through the thinking and doing aspects of assessment. This important tool will contain assessment experts and be a stretch for practitioners.

***Use of Existing Measurement Systems***

Typically HRD professionals begin an intervention assuming that everything they need has to be created from scratch. According to the *Results Assessment System*, the truth is exactly the opposite, particularly in the performance domain.

The *Results Assessment System* works under the assumption that if the intervention is targeted at real performance-related needs that are important, the chances are very good that someone in the organization is already measuring it. The closer assessment systems are to established organizational measurement systems, the more likely they are to last. This practical feature of the system should be received positively by both assessment experts and practitioners.

*Degree of Accuracy Required*

According to the *Results Assessment System*, the purpose of results assessment is to enable the organization to make sound decisions about interventions. The level of accuracy needed for organizational decisions is considerably less than that needed for research. The task is to find the "sweet spot" on the accuracy continuum so as to buy enough accuracy to slightly exceed management expectations needed to earn maximum credibility. If you buy more accuracy than that, you are wasting resources.

For example, consider an organization development program that is likely to return the typical financial return of 800% in a year or less (Swanson, 1998). Will it really matter to management if your measure is off by 50 percentage points, that is, you report 750% or 800%? Will you make any better decision about the program if you knew the return was actually 800%? If the answer is yes, then pay for that degree of accuracy. If the answer is no, accept less. Often, giving up a modest amount of accuracy will save considerable resources.

This concept of degree of accuracy will be discomforting to both assessment experts wanting "accuracy" and practitioners fearing "mistakes." In both cases, experience will likely be the mediator.

*Results Versus Formative Measures*

Evaluation professionals may assess inputs, processes, or outputs. For example, evaluators conduct formative evaluations during program pilot tests to assess effectiveness of a revised work process. While this is very valuable, the point is that evaluation models allow evaluators to skip measuring results! In fact, a popular 4-level evaluation scheme in clearly gives permission to practitioners to not measure results (Kirkpatrick, 1998).

To the organizations in which development is embedded, it is results that are paramount. All of the learning process and in-process perception measures are important, but are only important to organizations if they are *combined* with result measures. Thus, the *Results Assessment System* does not incorporate more formative measures because they are well handled elsewhere, not because they are not valued. Both assessment experts and practitioners should value the system's focus on the commonly overlooked domain of results.

**Conclusions**

The *Results Assessment System* is a fully developed assessment system for the Human Resource Development profession. It provides a theoretically sound alternative to the atheoretical tools and models that presently dominate professional practice. While the *Results Assessment System* is essentially a guide for practitioners, it holds promise as a useful research tool.

**References**


Alliger, G.M., & Janak, E.A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personal Psychology, 42*, 331-340.

Alliger, G., Tannenbaum, S.I., Bennett Jr., W., Traver, H., & Shotland, A. (1997). A Meta-Analysis of the Relations Among Training Criteria. *Personnel Psychology, 50*, 341-358.

Bassi, L. J., Gallagher, A. L., & Schroer0, E. (1996). *The ASTD training data book.* Alexandria, VA: ASTD Press.

Bassi, L., Benson, G. & Cheney, S. (1996). The top ten trends. *In Trends: Position Yourself for the Future,* Alexandria: ASTD Press, 2-16.

Dixon, N. M. (1990). The relationship between trainee responses on participation reaction forms and posttest scores. *Human Resource Development Quarterly, 1*, 129-137.

Holton, E.F. (1996). The flawed four-level evaluation model, 2nd edition. *Human Resource Development Quarterly, 7*(1), 5-29.

Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels.* San Francisco: Berrett-Koehler.

Micklewait, J. & Woolridge. A. (1996). *The witch doctors: Making sense of the management gurus.* New York: Random House.

Newstrom, J. (1996). Review: Evaluating training programs: The four levels. *Human Resource Development Quarterly. 6* (3). 317-320

Phillips, J. J. (1983). *Handbook for training evaluation and measurement methods.* Houston: Gulf.

Swanson, R. A. (1997). TADD Short! *Human Resource Development Quarterly. 8*(3) 193-195.

Swanson, R. A. (1998). Demonstrating the financial benefit of human resource development: Status of the theory and practice. *Human Resource Development Quarterly, 9*(3), 285-295.

Swanson, R. A. & Holton, E. F. (1999). *Results: How to assess performance, learning, and perceptions in organizations.* San Francisco: Berrett-Koehler.